ED 023 170 24 EA 001 661

By-Lindman, Erick L.
Net-Shift Analysis for Comparing Distributions of Test Scores.
California Univ., Los Angeles. Center for the Study of Evaluation of Instructional Programs.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Report No-WP-5
Bureau No-BR-6-1646
Pub Date Mar 68
Contract-OEC-4-6-061646-1909
Note-18p.
EDRS Price MF-$0.25 HC-$1.00
Descriptors-*Educational Testing, *Equated Scores, Program Evaluation, *Research Methodology, Statistical Analysis, *Testing, Testing Problems

A suggested technique for analyzing distributions of test scores compares distributions of scores made by groups of pupils on standard tests with distributions made by other groups of students on the same tests. By identifying the percents of student scores which must be shifted to an adjacent cell (interval) to make the two distributions exactly the same, the technique reveals changes in score distributions which may occur when different teaching methods are used. The technique also provides a more complete comparison between the distribution of scores made by a selected group of pupils and a norm group. (Author/JK)

BR 6-1646
PA 24

Working Paper No. 5

NET-SHIFT ANALYSIS FOR COMPARING DISTRIBUTIONS OF TEST SCORES

Erick L. Lindman

ED023170

Center FOR THE Study of Evaluation OF INSTRUCTIONAL PROGRAMS

University of California, Los Angeles, March 1968

EA 001 661

CO-DIRECTORS

Merlin C. Wittrock Erick L. Lindman

ASSOCIATE DIRECTORS

Marvin C. Alkin Frank Massey, Jr. C. Robert Pace

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds. Three major aspects of the program are

Instructional Variables - Research in this area will be concerned with identifying and evaluating the effects of instructional variables, and with the development of conceptual models, learning theory and theory of instruction. The research involves the experimental study of the effects of differences in instruction as they may interact with individual differences among students.

Contextual Variables - Research in this area will be concerned with measuring and evaluating differences in community and school environments and the interactions of both with instructional programs. It will also involve evaluating variations in student and teacher characteristics and administrative organization.

Criterion Measures - Research in this field is concerned with creating a new conceptualization of evaluation of instruction and in developing new instruments to evaluate knowledge acquired in school by measuring observable changes in cognitive, affective and physiological behavior. It will also involve evaluating the cost-effectiveness of instructional programs.

NET-SHIFT ANALYSIS FOR COMPARING DISTRIBUTIONS OF TEST SCORES

Erick L. Lindman

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

(Each paper printed as a Technical Report, Occasional Report, or Working Paper of the Center for the Study of Evaluation of Instructional Programs (UCLA) is reviewed prior to acceptance for publication. Standard journal review procedures are followed which include submission of comments to the Director by professional reviewers. This insures that the technical competence of the papers will be maintained at the high level set by the Center.)

CSEIP Working Paper No. 5, March 1968
University of California, Los Angeles

ABSTRACT

This paper suggests a technique for analyzing distributions of test scores. The technique is intended for comparing distributions of scores made by groups of pupils on standard tests with distributions made by other groups of students upon the same tests. Briefly, it does this by identifying the percents of student scores which must be shifted to an adjacent cell (interval) to make the two distributions exactly the same.

The technique is intended to reveal changes in score distributions which may occur when different teaching methods are used. For example, a new remedial program might cause a shift of low scores toward the mean without altering the distribution of scores above the mean. The technique also provides a more complete comparison between the distribution of scores made by a selected group of pupils and a norm group.

NET-SHIFT ANALYSIS FOR COMPARING DISTRIBUTIONS OF TEST SCORES

Evaluations of educational programs usually require comparisons of test score distributions. Three types of comparisons are common:

1. Comparison between the distributions of scores made on a standardized test by a selected group of children and a national or state norm distribution of scores for the same test.
2. Comparison between the distributions of scores made on the same test by successive grade level groups in a school. For example, a city school system may wish to compare the distribution of scores made by third grade children on a reading test with the distribution of scores made on the same test by former third grade groups.
3. Comparison between the distributions of scores made by the same group of students on different tests. For example, comparison between arithmetic and reading scores for fourth grade children in the same school may provide an indication of relative effectiveness of the teaching of arithmetic and reading.

Note that in 1 and 2 above, comparisons are between distributions of scores made by different groups of pupils; while in 3 above, the comparisons are between distributions of scores made

by the same pupils on different tests. Only in the latter case is correlational analysis possible.

In comparing distributions of test scores, often only measures of central tendency are considered. One frequently hears the statement, "Our fifth graders are above the national average in reading." Or perhaps the statement is a little more precise: "The average fifth grader in our school scored above the national average score for fifth graders." In either case the percent of "our fifth graders" that scored in the lowest 10 percent of the national norm distribution and the percent that scored in the highest 10 percent, for example, are not revealed. Such information about the entire distribution of scores is essential for evaluating the reading achievement of "our fifth graders."

The test score distribution comparison procedure proposed in this paper seeks to accomplish two basic purposes:

1. To compare the distribution of test scores for a group of students to a corresponding national or state norm distribution in such a way that the entire distributions are compared.
2. If one group of students has a higher average score than another, to locate the points along the entire distribution that account for the difference in the average scores. A shift in average score may reflect shifts among the low scores, the middle scores or the high scores in unequal amounts. This type of analysis is needed to compare successive score distributions before and after teaching methods have been changed to determine if the new method

tends to increase or decrease scores in one part of the distribution more or less than in other parts.

To use the proposed distribution comparison procedure, it is first necessary to convert the reference distribution (which may be a national, or school district state norm) to some standard distribution such as deciles or stanines. The decile or stanine intervals of the reference distribution (in raw scores) provide the intervals for all distributions of scores of study groups. In Exhibits I and II, the row labeled Raw Score Ranges contains in each cell the raw score interval corresponding to the percent shown above it. By this process distributions of scores of study groups can be quickly compared with the reference distribution. For example, if the decile distribution is used, the percent of the scores of the study group that is in the upper 10 percent or upper 20 percent of the norm distribution is indicated in the appropriate cell. Similarly, if the stanine distribution is used, the percent of the study group that scored in the upper 3 percent or upper 11 percent is indicated.

Ordinarily, in using this procedure, two study groups, A and B, are compared with the reference distribution and with each other. In a typical case, distribution A might be the third grade reading scores made on a test last year and distribution B might be this year's third grade scores on the same test. We are interested in comparing both study group distributions A and B with the national norm distribution and with each other.

Exhibit I shows the computations when raw scores of the reference distribution are converted to "deciles." The row labeled Raw

Score Ranges shows the ranges of scores in each cell containing successive tenths of the reference distribution. These raw score intervals are then used to determine the percents of the A and B distributions in each cell.

By comparing the percents of scores of the A or B distribution in each cell with the corresponding percent for the reference distribution, one can quickly answer such questions as: What percent of the A group scored above the norm median? What percent of the B group scored in the lowest 10 percent of the norm group? Answers to such questions make comparisons with the reference group more meaningful.

Exhibits I and II show how these comparisons are made. The row called "Raw Score Ranges" shows the range for each interval. Immediately below this range is the percent of the reference distribution in each cell. Since distributions A and B are based upon the intervals established by the reference distribution, the percents shown in these rows (A_1, A_2 ... and B_1, B_2 ...) are directly comparable with the percents of the reference distribution in the corresponding cell.

However, the proposed comparison procedure is designed primarily to compare the score distributions of two study groups, A and B, with each other. To accomplish this, row C is obtained by subtracting percents entered in the corresponding cells of rows A and B. These differences will, of course, total zero as indicated in the right-hand column; that is, ΣC = zero.

The next step is to enter in each cell of row X the cumulative totals computed from row C. In the first cell of row X,

the amount X_1 equals C_1. In the second cell of row X, the amount X_2 equals X_1 plus C_2. Similarly, X_3 equals X_2 plus C_3. This procedure is continued until amounts are entered in the last cell of row X. Note that the amount entered in the last cell of row X immediately to the left of the total column always will be zero.

The positive percents shown in row X may be interpreted as the percent of scores in row A, which must be shifted to the next cell on the right to make all entries in row A equal to corresponding entries in row B. Negative percents in row X are interpreted as the percents of scores in row A, which must be shifted to the left <u>from</u> the cell immediately to its right in order to make the distribution of percents in row A exactly the same as those in row B. The total of row X indicates the aggregate net shift necessary to make the percent in each cell of row A exactly equal to the corresponding percent in row B. Thus, row X indicates how much rows A and B differ <u>and in which cells these differences occur</u>.

Interpretation of row X as the number of "shift units" which must be applied to the entry in each cell of row A to make it equal to the entry in the corresponding cell of row B provides a useful way to compare distributions. Consider a hypothetical case in which row A in Exhibit III represents the distribution of reading scores before a remedial program was introduced and row B represents the distribution after the remedial program was introduced. Row C shows the cell differences and row X the accumulative totals of row C.

How can the change which occurred in the distribution of scores shown in Exhibit III be described? In familiar terms, the

median is unchanged, the mean has increased, and the variance has decreased; but this hardly tells the story. Nor are the cell differences shown in row C very helpful; they seem to indicate that there were 5 percent shifts in the two lowest decile cells, with corresponding 5 percent losses in the next two higher cells.

Row X is much more informative. There has been a net shift of 20 "shift units," representing a shift of 5 percent of the scores in distribution A from the lowest to the next higher decile cell--a shift of 10 percent from the second to the third decile cell and a shift of 5 percent of the scores from the third to the fourth decile cell. A shift of one "shift unit" means that one percent of the scores has shifted to the adjacent cell on the right. Similarly, a loss of one "shift unit" (or a negative "shift unit") means that one percent of the scores has shifted to the left from the adjacent cell on the right. Note that a shift of 2 percent of the scores to the next adjacent cell on the right or a shift of one percent of the scores to the second cell on the right has the same effect upon the aggregate net shift of the distribution.

Utilizing the shift units to describe the difference in two distributions makes it possible not only to describe the total amount of the difference, but also to describe where throughout the distribution the differences have occurred. Note that a gain of 20 shift units does not mean that 20 identifiable individuals shifted from one cell to the next higher cell. The 20 is a percent and may represent any number of individuals. Since the N in Exhibit III is more than 5,000, 20 percent represents more than

1,000 scores. Moreover, some hypothetical scores may have moved more than one cell to the right and some may have moved to the left. Actually, since different individuals are in the two distributions, there is no way to trace a specific score. The net shift merely describes the difference between the two distributions much as if comparisons were made between their means and standard deviations.

The total of row X is closely related to the difference of the means of distributions A and B. When scores are recorded as "stanines," the sum of row X divided by 100 equals the difference of the means of rows A and B in stanine units. In this case the procedure distributes the difference of the two means among the cells so that one can tell if the observed difference is due to changes concentrated at one end or the other of the distribution.

When scores are recorded in "deciles," the total of row X divided by 100 is not equal to the difference of the means of rows A and B, because the score differences between the decile intervals are not equal. In this case, the total of row X is approximately proportional to the difference of the means of rows A and B. In either case the important point is that the aggregate shift or the difference in the means of rows A and B can be divided into components located at different points along the distribution.

Although inspection of row X gives a general indication of the extent to which gains or losses have occurred at one end of the distribution or at the other, a more precise measure may be useful. For this purpose, row Y is computed by entering the

cumulative totals from row X in the corresponding cells of row Y. Row Y is derived from row X, precisely as row X was derived from row C. For example, Y_3 equals Y_2 plus X_3.

It will be noted that for the "decile" scores, the sum of row Y is equal to $9X_1$ plus $8X_2$ plus $7X_3 \ldots +1X_9$. Thus, the X's are weighted in a descending order from left to right, giving more weight to low scores.

By comparing the sum of row Y with the sum of row X, it is possible to obtain more precise indicators of the location within the distribution where gains or losses have occurred. The sums shown on the lower part of Exhibits I and II are for this purpose.

The sum of the X's (ΣX) is a measure of the amount by which the average of distribution B exceeds A. A negative total indicates that the average of distribution of A exceeds B. ΣX is designated as the aggregate shift.

The weighted low-score shift, $(1/10)\ \Sigma Y$, indicates whether the aggregate shift occurred mainly among the high or low scores. If this index equals one-half of the aggregate shift, high and low score changes contribute equally to the overall difference between distributions A and B. If the weighted low-score shift is greater than one-half of the aggregate shift, more of the shift occurred among the low scores than among the high scores.

The weighted high-score shift is obtained by subtracting the weighted low-score shift from ΣX. A relatively large, weighted high-score shift (more than one-half of the aggregate shift) indicates that most of the shift occurred among the high scores.

Although the weighted high-score shift is obtained by subtracting the weighted low-score shift from ΣX, the weighted high-score shift is a weighted sum of the X's in which greater weights are given to the high scores. This can be seen from the following relationships:

$$10\Sigma X = 10X_1 + 10X_2 + 10X_3 \ldots 10X_8 + 10X_9$$
$$\Sigma Y = 9X_1 + 8X_2 + 7X_3 \ldots 2X_8 + X_9$$
$$10\Sigma X - \Sigma Y = X_1 + 2X_2 + 3X_3 \ldots 8X_8 + 9X_9$$

Thus, by introducing a factor of 10 before the subtraction is made, it is clear that ΣY and 10ΣX-ΣY are weighted sums of the X's in which the weightings are reversed. One gives greater weightings to low scores on the left of the distribution and the other gives greater weightings to high scores on the right of the distribution. For this reason they are indicators of the extent to which gains or losses have occurred primarily among the low scores or high scores.

This type of analysis becomes increasingly important as we interpret the meaning of equal educational opportunity and seek to devote more educational resources to slow learners. We need to know if an instructional program is reducing or increasing the variation of test score distributions and if it is especially effective at one end of the distribution.

The net-shift analysis of test-score distributions before and after an instructional treatment provides essential information concerning its effect upon the distribution of student scores. In some cases it may be appropriate to use the normalized pretest distribution as the reference distribution. In

such an analysis, the percents shown in row A would be equal to the corresponding percents of the reference distribution. The procedure might be useful if the study group differs greatly from a national, state, or local norm group.

Summary

The basic technique suggested in this paper differs from customary procedures for comparing distributions of test scores in two respects. First, a reference distribution is used in place of norms expressed only by measures of central tendency and variability. Second, the intervals of the reference (or norm) distribution are used to group scores of the distributions being studied.

In some respects the procedure is similar to the Chi square analysis since one distribution may be considered to be the expected, and the other the actual distribution. However, instead of squaring the differences between the expected and actual number of scores in each cell, the differences are accumulated to determine the percent of scores which must be shifted to the next higher cell (or, if negative, the next lower cell) to make the distributions exactly equal.

The net-shift analysis preserves the signs which indicate the direction of the shift. This information, lost in the Chi square analysis, is essential, especially if some shifts are positive and some negative.

Moreover, in the net-shift analysis, shifts from a cell at one end of the distribution to a cell at the other end are weighted more heavily than shifts between adjacent cells. In

the Chi square analysis there is no distinction between these types of "shifts." In comparing distributions of test scores, it is obvious that more change has occurred if 10 percent of the scores shift from the lowest to the highest quartile than if 10 percent of the scores shift from the first to the second quartile. In this respect, the net-shift analysis provides a more complete description of the differences between two distributions.

The weighted low-score shift and the weighted high-score shift are intended to provide measures of the extent to which gains or losses tend to be concentrated at one end or the other of the distribution. In most cases, examination of row X will be more informative than the weighted low-score or high-score shift. However, if many distributions are under study and if programs intended especially for slow learners or for the gifted have been used, the weighted low-score and high-score shifts may be useful for comparison purposes.

EXHIBIT I

Net-Shift Analyses of Test Score Distributions (Decile Form)

Work Sheet A
Lindman

For School _____ Date _____

Distribution A _____ N = ( )

Distribution B _____ N = ( )

Reference Distribution _____

Decile	Low 10%	2nd 10%	3rd 10%	4th 10%	5th 10%	6th 10%	7th 10%	8th 10%	9th 10%	High 10%	Total
Reference Distribution											
Raw Score Ranges											XXX
Pct. in each cell	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%

Study Distributions

Pct. of scores	A	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	100%
in each cell	B	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	100%
C = A-B	C	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	0
Cum. of C	X	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	0	$\Sigma X=$
Cum. of X	Y	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	XXX	$\Sigma Y=$

Indices: ΣX _____ Aggregate shift: (1/10) ΣY _____ Weighted low-score shift

ΣX - (1/10) ΣY _____ Weighted high-score shift

EXHIBIT II

Net-Shift Analyses of Test Score Distributions (Stanine Form)

Work Sheet B
Lindman

For School _____ Date _____

Distribution A _____ N = ( )

Distribution B _____ N = ( )

Reference Distribution _____

Stanines	1	2	3	4	5	6	7	8	9	Total
Reference Distribution										
Raw Score Ranges										XXX
Pct. in each cell	4%	7%	12%	17%	20%	17%	12%	7%	4%	100%

Study Distributions

Pct. of scores in each cell	A	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	100%
	B	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	100%
C= A-B	C	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	0
Cum. of C	X	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	0	ΣX=
Cum. of X	Y	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	XXX	ΣY=

Indices: ΣX _____ Aggregate shift; (1/9) ΣY _____ Weighted low-score shift

ΣX - (1/9) ΣY _____ Weighted high-score shift

EXHIBIT III

Net-Shift Analyses of Test Score Distributions (Decile Form)

Work Sheet A
Lindman

For School Garfield School Date April, 1968

Distribution A Third Grade (1967) Before Remedial Program N = (5,200)

Distribution B Third Grade (1968) After Remedial Program N = (5,300)

Reference Distribution National Reading Test - Cal. State Norm

Decile		Low 10%	2nd 10%	3rd 10%	4th 10%	5th 10%	6th 10%	7th 10%	8th 10%	9th 10%	High 10%	Total
Reference Distribution												
Raw Score Ranges												XXX
Pct. in each cell		10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%
Study Distributions												
Pct. of scores	A	15	15	10	10	10	10	10	10	5	5	100%
in each cell	B	10	10	15	15	10	10	10	10	5	5	100%
C = A-B	C	5	5	-5	-5	0	0	0	0	0	0	0
Cum. of C	X	5	10	5	0	0	0	0	0	0	0	ΣX= 20
Cum. of X	Y	5	15	20	20	20	20	20	20	20	XXX	ΣY=160

Indices: ΣX 20 Aggregate shift; (1/10) ΣY 16 Weighted low-score shift

ΣX - (1/10)ΣY 4 Weighted high-score shift